K.R.K. Srivatsava
MT18054

Exercise 3.15

We have,

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots$$

$$= \gamma^0 R_{t+0+1} + \gamma^1 R_{t+1+1} + \gamma^2 R_{t+2+1} + \cdots$$

$$= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \longrightarrow (1)$$

if the reward is a constant +1

we have

$$G_t = \sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma} \longrightarrow (2)$$

So for a constant reward 'c'

$$G_t = \sum_{k=0}^{\infty} \gamma^k = c\left[\frac{1}{1-\gamma}\right]$$

We Know have the state value function

$$v_\pi(s) = E_\pi\left[G_t \mid S_t = s\right]$$

$$= E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s\right]$$

if a constant c is added to the reward

$$V'_\pi(s) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + c) \,\middle|\, S_t = s\right]$$

$$= \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,\middle|\, S_t = s\right] + c\,\mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k \,\middle|\, S_t = s\right]$$

$$= V_\pi(s) + c\left[\frac{1}{1-\gamma}\right]$$

$V_c = \frac{c}{1-\gamma}$; V_c is a constant state value term added upon adding c to each reward

---

Exercise 3.16

In an episodic task, addition of c will result in the following modification of the above term

$$\sum_{k=0}^{n} \gamma^k = \frac{1-\gamma^n}{1-\gamma}$$

$$\therefore V'_\pi(s) = V_\pi(s) + c\left[\frac{1-\gamma^n}{1-\gamma}\right]$$

Express v_* in terms of q_*

Optimal state-value function

$$v_*(s) = \max_{\pi} v_{\pi}(s)$$

Optimal ~~state~~ action-value function

$$q_*(s,a) = \max_{\pi} q_{\pi}(s,a)$$

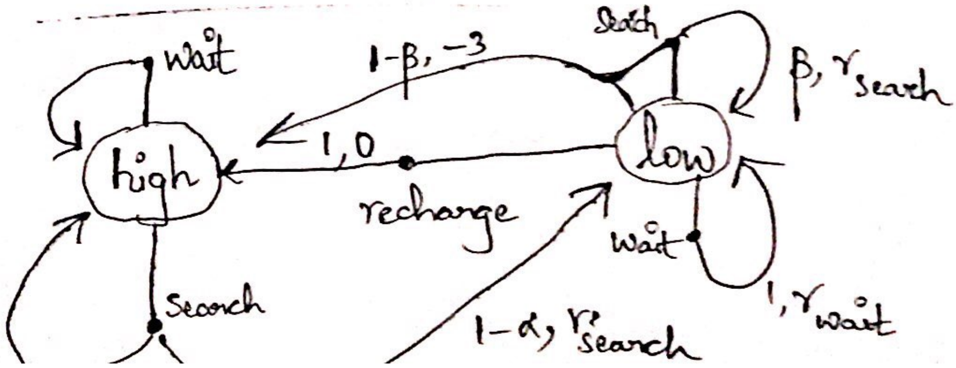
$$v_*(s) = \max_{a \in A(s)} q_{\pi_*}(s,a)$$

$$= \max_{a} E_{\pi}\left[G_t \mid S_t = s ; A_t = a\right]$$

$$= \max_{a} E_{\pi}\left[R_{t+1} + \gamma v_*(S_{t+1}) \mid S_t = s ; A_t = a\right]$$

$$= \max_{a} \sum p(s', r \mid s,a)\left[r + \gamma v_*(s')\right]$$

3.4



For the above MDP, we have,

s	a	s'	r	$p(s', r \mid s, a)$
high	search	high	r_{search}	α
high	search	low	r_{search}	$1-\alpha$
high	wait	high	r_{wait}	1
low	wait	low	r_{wait}	1
low	search	low	r_{search}	β
low	search	high	-3	$1-\beta$
low	~~wait~~ recharge	high	0	1